PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-293317
(43)Date of publication of application : 20.10.2000

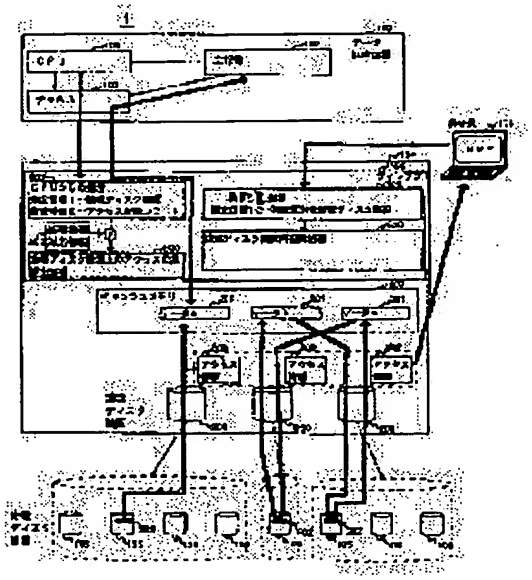
(51)Int.Cl. G06F 3/06
G06F 12/16
G11B 19/02
G11B 20/12

(21)Application number : 2000-066061
(22)Date of filing : 08.04.1996

(71)Applicant : HITACHI LTD
(72)Inventor : YAMAMOTO YASUTOMO
YAMAMOTO AKIRA
SATO TAKAO

(54) STORAGE CONTROLLER

(57)Abstract:
PROBLEM TO BE SOLVED: To improve the access performance in sequential access, etc., by performing rearrangement to a physical storage device in units of logical storage drives and successively storing data on the physical storage device.
SOLUTION: A service engineer refers to access information 500 presented by an SVP 111 to examine the rearrangement of the logical disk drives 200. Consequently, when there is a logical disk 200 decided to be rearranged, a rearrangement indication 620 is sent to the storage controller 104. A director 106 receives the indication 620 and performs a logical disk rearranging process 630 between two specified logical disk drives 200. At this time, logical-physical correspondence information 300 is used to transfer data in one-process units to be rearranged from the physical disk drive 105 to a cache memory 107. Then the data 201 in the process units on the cache memory 107 are repeatedly written to the physical disk drive 105 as a rearrangement destination and the correspondence information 300 is updated after the writing is completed.



LEGAL STATUS

[Date of request for examination] 27.03.2003

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]
[Claim 1] The memory control unit characterized by to have a logical-memory equipment relocation means store data in the physical store of a relocation place continuously while a data processor rearranges said logical-memory equipment to said physical store based on the index defined beforehand in the memory control unit which arranges the logical memory equipment which performs direct access to the physical store which actually memorizes data, and controls the data transfer between said data processors and said physical stores.
[Claim 2] In the memory control unit with which a data processor controls the data transfer between matching, said data processor, and said physical store for the physical store which actually remembers data to be logical memory equipment which performs direct access An access information extraction means to extract the access information to the logical memory equipment of a data processor as an index during employment of control of said data transfer, The memory control unit characterized by having a logical memory equipment relocation means to store data in the physical store of a relocation place continuously while rearranging said logical memory equipment to said physical store based on said index.
[Claim 3] The memory control unit with which said access information is characterized by including the access frequency information from said data processor to said logical memory equipment in a memory control unit according to claim 2.
[Claim 4] The memory control unit with which said access information is characterized by including the access pattern information from said data processor to said logical memory equipment in a memory control unit according to claim 2 or 3.
[Claim 5] The memory control unit characterized by said index being the dependability for which said logical memory equipment is asked in a memory control unit according to claim 1.
[Claim 6] The memory control unit characterized by providing an index presentation means to show said index to either of claim 1 to claims 5 in the memory control unit of a publication at a customer engineer, and a relocation directions reception means to receive the relocation directions from a customer engineer.
[Claim 7] The memory control unit characterized by providing a relocation directions reception means to receive the relocation directions from a data processor in a memory control unit given in either of claim 1 to claims 5.
[Claim 8] The memory control unit characterized by providing a relocation necessity decision means to determine the necessity of relocation as either of claim 1 to claims 5 in the memory control unit of a publication based on said index.
[Claim 9] When the logical memory equipment under relocation has access from a data processor in the memory control unit of a publication at either of claim 1 to claims 8, The completion field of relocation and relocation incomplete field of logical memory equipment under relocation are identified. The memory control unit characterized by providing further the access location change means which will be made to access the logical memory equipment of a relocation place if said access location is said completion field of relocation, and will be made to access the logical memory equipment concerned if said access location is said relocation incomplete field.

[Claim 10] It is the store system characterized by for said memory control unit to acquire the access situation by said data processor in the store system which has the memory control unit which controls the data transfer between two or more physical stores for holding the data of the logical store which a data processor recognizes for a data access, and said two or more physical units and data processor, and to move the data of said logical store to the 2nd physical store from the 1st physical store based on said access situation.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention] This invention relates to the memory control unit which can improve the dependability of the memory control unit which can improve the access engine performance with the case of a sequential access, or a random accelerator even when a hit ratio is low, and data in more detail about a memory control unit. Especially this invention is useful to the store subsystem constituted by the highly efficient disk unit, its highly efficient disk unit, and disk controller of the disk array sense, and the information processing system constituted by the store subsystem and data processor.
[0002]
[Description of the Prior Art] The paper "D. Patterson, G.gibson, and R.H.Kartz;A Case for Redundant Arrays of Inexpensive Disks (RAID), ACM SIGMOD Conference, Chicago and IL, (June 1988) pp.109-116" announced in the "ACM SIGMOD" meeting held in University of Illinois of Chicago is indicating the technique about the data arrangement on a disk array.
[0003] Moreover, in JP,7-84732,A, the technique of using a part of disk unit like a disk cache is indicated. It specifically divides into the temporary field which stores data for a disk unit temporarily, and the field which finally writes in data, and duplex writing of the updating data is once carried out to a temporary field, without generating parity, they carry out parity generation asynchronous, and are written in the last field.
[0004] On the other hand, the technique of changing the RAID level holding data dynamically by the difference in access frequency is indicated by the electric information American Communications Association technical research report "DE 95-68 (others [Motegi]: the performance evaluation at the time of disk failure of the disk array using Hot Mirroring, December, 1995, electric information American Communications Association technical report Vol.95-No.407, pp.19-24)." By dividing a disk unit into the part of RAID1 configuration, and the part of RAID5 configuration, and specifically changing the storing location of data dynamically so that data with light access may be preferentially stored in the part of RAID1 configuration, the data with high access frequency are stored in the part of RAID1 configuration, and what has low access frequency can be stored in the part of RAID5 configuration. According to this technique, it is possible to make the physical disk equipment with which the physical disk equipment with which memory capacity differs differs from RAID level intermingled within a store subsystem, and the data in logical disk equipment can be stored in the physical disk equipment of arbitration based on indexes, such as that access frequency, access pattern, etc. Moreover, a storing location can also be dynamically changed so that data with high access frequency may be stored in more nearly high-speed physical disk equipment. In addition, the disk array of RAID1 writes the duplicate in the subdisk unit called a mirror to the write-in data from a data processor, and secures the dependability of data. Since redundancy data are the duplicate of the original data, the overhead of redundancy data origination is small and the access engine performance is good. However, the utilization ratio of physical storage is as low as 50%. On the other hand, the disk array of RAID5 creates the redundancy data called parity to two or more write-in data from a data processor. Since the data before updating and the parity before updating need to be led,

the overhead of redundancy data origination is large to parity creation time, and the access engine performance is bad to it. However, in order to create one parity to two or more data, the utilization ratio of storage is high compared with RAID1.
[0005]
[Problem(s) to be Solved by the Invention] With the above-mentioned conventional technique, in order to change the storing location of data by the data unit to access, on the logical disk equipment with which a data processor performs direct access, data [****] will be discontinuous on the physical disk equipment which actually memorizes data. For this reason, in the case of the sequential access which carries out read/write of a series of data, in fact, two or more data are gathered and there is a trouble which stops being able to carry out read/write and causes access performance degradation.
[0006] Many of data which the access pattern moved with the random accelerator at the part of RAID1 configuration when a hit ratio was low in order to write light data in the part of RAIDwhich moved data judged that access frequency is low on the other hand at every light with conventional technique of the above-mentioned report "DE 95-68" to part of RAID5 from part of RAID1 configuration configuration, and was vacant 1 configuration will be again returned to the part of RAID5 configuration. For this reason, when a hit ratio is low, the improvement in the access engine performance cannot be expected, but has the trouble that the overhead of processing which moves data conversely causes access performance degradation.
[0007] Moreover, with the above-mentioned conventional technique, there is a trouble which is not taken into consideration at all about improvement in the dependability of data.
[0008] Then, with the case of a sequential access, or a random accelerator, the 1st purpose of this invention is to offer the memory control unit which can improve the access engine performance, even when a hit ratio is low. Moreover, the 2nd purpose of this invention is to offer the memory control unit which can improve the dependability of data.
[0009]
[Means for Solving the Problem] In the 1st viewpoint, a data processor arranges this invention to the physical store which actually memorizes data for the logical memory equipment which performs direct access. In the memory control unit which controls the data transfer between said data processors and said physical stores While rearranging said logical memory equipment to said physical store based on the index defined beforehand, the memory control unit characterized by having a logical memory equipment relocation means to store data in the physical store of a relocation place continuously is offered. In the memory control unit by the 1st viewpoint of the above, the storing location of data is not changed by the data unit to access, relocation to physical storage is performed for logical memory equipment as a unit, and data are continuously stored in the physical storage of a relocation place. Therefore, also in the case of a sequential access, the access engine performance can be improved. Moreover, the storing location of data is not changed at every light, but since said relocation is performed based on the index defined beforehand, even when a hit ratio is low, the access engine performance can be improved with a random accelerator.
[0010] In the 2nd viewpoint, this invention matches the physical store which actually remembers data to be logical memory equipment with which a data processor performs direct access. In the memory control unit which controls the data transfer between said data processors and said physical stores An access information extraction means to extract the access information to the logical memory equipment of a data processor as an index during employment of control of said data transfer, While rearranging said logical memory equipment to said physical store based on said index, the memory control unit characterized by having a logical memory equipment relocation means to store data in the physical store of a relocation place continuously is offered. In the memory control unit by the 2nd viewpoint of the above, the storing location of data is not changed by the data unit to access, relocation to physical storage is performed for logical memory equipment as a unit, and data are continuously stored in the physical storage of a relocation place. Therefore, also in the case of a sequential access, the access engine performance can be improved. Moreover, since the storing location of data is not changed at every light, but access information is extracted and said relocation is performed, using it

statistically, even when a hit ratio is low, the access engine performance can be improved with a random accelerator.

[0011] In the 3rd viewpoint, this invention offers the memory control unit characterized by said access information including the access frequency information from said data processor to said logical memory equipment in the memory control unit of the above-mentioned configuration. In the memory control unit by the 3rd viewpoint of the above, logical memory equipment with high access frequency is rearrangeable to more nearly high-speed physical storage. Therefore, the access engine performance can be improved.

[0012] In the 4th viewpoint, this invention offers the memory control unit characterized by said access information including the access pattern information from said data processor to said logical memory equipment in the memory control unit of the above-mentioned configuration. In the memory control unit by the 4th viewpoint of the above, logical memory equipment with the high ratio of a sequential access is rearrangeable to physical storage with the more high sequential access engine performance. Therefore, the access engine performance can be improved.

[0013] In the 5th viewpoint, this invention offers the memory control unit characterized by said index being the dependability for which said logical memory equipment is asked in the memory control unit of the above-mentioned configuration. In the memory control unit by the 5th viewpoint of the above, that it is reliable can rearrange the logical memory equipment called for to more reliable physical storage. Therefore, the dependability of data can be improved.

[0014] In the 6th viewpoint, this invention offers the memory control unit characterized by providing an index presentation means to show a customer engineer said index, and a relocation directions reception means to receive the relocation directions from a customer engineer in the memory control unit of the above-mentioned configuration. In the memory control unit by the 6th viewpoint of the above, since a customer engineer can input relocation directions, said relocation can be performed very flexibly.

[0015] In the 7th viewpoint, this invention offers the memory control unit characterized by providing a relocation directions reception means to receive the relocation directions from a data processor in the memory control unit of the above-mentioned configuration. In the memory control unit by the 7th viewpoint of the above, since a data processor can input relocation directions, a customer engineer can perform said relocation under the advanced conditions which cannot be judged.

[0016] In the 8th viewpoint, this invention offers the memory control unit characterized by providing a relocation necessity decision means to determine the necessity of relocation based on said index in the memory control unit of the above-mentioned configuration. In the memory control unit by the 8th viewpoint of the above, in order for a memory control unit to make a self-decision of the relocation directions, it is not necessary to apply a burden to a customer engineer or a data processor.

[0017] When this invention has access from a data processor in the logical memory equipment under relocation in the memory control unit of the above-mentioned configuration in the 9th viewpoint, The completion field of relocation and relocation incomplete field of logical memory equipment under relocation are identified. If said access location is said completion field of relocation, the logical memory equipment of a relocation place will be made to access, and if said access location is said relocation incomplete field, the memory control unit characterized by providing the access location change means made to access the logical memory equipment concerned will be offered. In the memory control unit by the 9th viewpoint of the above, since the completion field of relocation and relocation incomplete field of logical memory equipment under relocation are identified and the access location from a data processor is changed, while employing the data transfer between a data processor and a physical store, it is rearrangeable.

[0018] Two or more physical stores for this invention to hold the data of the logical store which a data processor recognizes for a data access in the 10th viewpoint, In the storage system which has the memory control unit which controls the data transfer between said two or more physical units and data processors said memory control unit The access situation by said data processor is acquired, and the storage process defined system characterized by moving the data

of said logical store to the 2nd physical store from the 1st physical store based on said access situation is offered. In the memory control unit by the 10th viewpoint of the above, since the data of logical memory equipment are moved to the 1st to 2nd physical store according to the access situation by the data processor, the access engine performance can be improved.

[0019]

[Embodiment of the Invention] Hereafter, the operation gestalt of this invention is explained. In addition, thereby, this invention is not limited.

[0020] - 1st operation gestalt - The 1st operation gestalt extracts the access information of each logical disk equipment with a memory control unit, and shows it to a customer engineer through SVP (service processor), and the relocation directions of a customer engineer based on this access information perform relocation to the physical disk equipment of logical disk equipment.

[0021] Drawing 1 is the block diagram of the information processing system containing the memory control unit concerning the 1st operation gestalt of this invention. This information processing system 1 has come to connect a data processor 100, a memory control unit 104, one or more physical disk equipments 105, and SVP111.

[0022] Said data processor 100 has CPU101, the primary storage 102, and the channel 103.

[0023] Said memory control unit 104 has one or more directors 106, cache memory 107, directory 108, nonvolatile memory 109, the nonvolatile memory management information 110, the information 300 corresponding to logic physics, the logical disk equipment information 400, and access information 500. Said director 106 performs data transfer between the channel 103 of a data processor 100, the data transfer between physical disk equipment 105 and the channel 103 of a data processor 100, and said cache memory 107, and data transfer between said cache memory 107 and physical disk equipment 105. Data with the high access frequency in physical disk equipment 105 are loaded to said cache memory 107. Said director 106 performs this load processing. The examples of the data to load are the data for access of CPU101 of a data processor 100, this data for access, data with the near storing location on physical disk equipment 105, etc. Said directory 108 stores the management information of said cache memory 107. Said nonvolatile memory 109 loads data with the high access frequency in physical disk equipment 105 like said cache memory 107. Said nonvolatile memory management information 110 stores the management information of said nonvolatile memory 109. Said information 300 corresponding to logic physics is information which shows the logical disk equipment (200 of drawing 2) arranged at the location and each physical disk equipment 105 on the physical disk equipment 105 with which each logical disk equipment (200 of drawing 2) is arranged. Calculation of the storing field on the physical disk equipment 105 of the data for access of CPU101 of a data processor 100 etc. is performed using this information. Said logical disk equipment information 400 shows conditions, such as access propriety of each logical disk equipment (200 of drawing 2). Said access information 500 is the information on the access frequency of each logical disk equipment (200 of drawing 2), an access pattern, etc.

[0024] The information 300 and the logical disk information 400 corresponding to logic physics are recorded on a non-volatilized medium, in order to prevent disappearance by power off etc.

[0025] Said physical disk equipment 105 consists of a medium which records data, and equipment which write the recorded data.

[0026] Said SVP111 receives an input of the presentation to the customer engineer of access information 500 and the relocation directions 620 from a customer engineer. Moreover, presentation to customer engineers, such as a fault condition of the dispatch of directions and information processing system 1 to the information processing system 1 from a customer engineer, is performed.

[0027] Drawing 2 is drawing showing the relation of logical disk equipment 200 and physical disk equipment 105. Logical disk equipment 200 is an apparent disk unit in which CPU101 of a data processor 100 carries out direct access, and corresponds with the physical disk equipment 105 with which the data for access are actually stored. The data on logical disk equipment 200 are continuously arranged on physical disk equipment 105 in consideration of the sequential access. When the physical disk equipment 105 with which the data of logical disk equipment 200 are

arranged is a disk array configuration, this logical disk equipment 200 corresponds with two or more physical disk equipments 105. Moreover, the capacity of physical disk equipment 105 is larger than logical disk equipment 200, and when the data of two or more logical disk equipments can be stored in one physical disk equipment 105, this physical disk equipment 105 corresponds with two or more logical disk equipments 200. Correspondence of this logical disk equipment 200 and physical disk equipment 105 is managed for said information 300 corresponding to logic physics. For example, when CPU101 of a data processor 100 leads the data 201 of logical disk equipment 200, based on the information 300 corresponding to logic physics, it asks for the physical disk equipment 105 corresponding to logical disk equipment 200 with a memory control unit 104, and asks for the data storage location 202 in the field of the physical disk equipment 105, and data transfer is performed.

[0028] Drawing 3 is drawing showing the information 300 corresponding to logic physics. The information 300 corresponding to logic physics consists of logical disk configuration information 310 and physical disk configuration information 320. Said logical disk configuration information 310 is the information about the field on the physical disk equipment 105 with which each logical disk equipment 200 is arranged, and when asking for the physical disk equipment 105 which corresponds from logical disk equipment 200, it is used. On the other hand, said physical disk configuration information 320 is the information about the logical disk equipment 200 arranged at each physical disk equipment 105, and when asking for the logical disk equipment 200 which corresponds from physical disk equipment 105, it is used.

[0029] As for said logical disk configuration information 310, only the number of logical disk equipment 200 has the group of the physical disk device group 311, the RAID configuration 312, and a starting position 313. Said physical disk device group 311 is information which shows the physical disk equipment 105 with which the logical disk equipment 200 concerned is arranged. Said RAID configuration 312 shows said physical disk device group's 311 RAID level. Said starting position 313 shows the head location where the logical disk equipment 200 concerned is arranged on physical disk equipment 105.

[0030] As for said physical disk configuration information 320, only the number of physical disk equipment 105 has the logical disk device group 321. Said logical disk device group 321 shows the logical disk equipment 200 arranged at the physical disk equipment 105 concerned.

[0031] Drawing 4 is drawing showing the logical disk information 400. As for the logical disk information 400, only the number of logical disk equipment 200 has the logical disk condition 401 and the completion pointer 402 of relocation. Said logical disk condition 401 expresses the condition of the logical disk equipments 200 "normal", "lock out", "under", etc. ["under a format and relocation"] Said completion pointer 402 of relocation is effective information only while said logical disk condition 401 "is rearranging", and the head location of the field which has not yet finished relocation processing is shown, the next location 200, i.e., logical disk equipment concerned, of the field which has completed relocation processing of the logical disk equipment 200 concerned. In access to the field before the completion pointer 402 of relocation, it must access to the physical disk equipment 105 after relocation at the time of the data access "under relocation." On the other hand, in access to the field after completion pointer of relocation 402, it must access to the physical disk equipment 105 before relocation.

[0032] Drawing 5 expresses access information 500. As for access information 500, only the number of logical disk equipment 200 has the access frequency information 501 and the access pattern information 502. Refer to this access information 500 for a memory control unit 104, a data processor 100, or SVP111. Said access frequency information 501 manages the count of access to the logical disk equipment 200 concerned per unit time amount. This access frequency information 501 is used as an index which asks for the high thing or the low thing of access frequency in each logical disk equipment 200. Said access pattern information 502 manages the rate of the sequential access to the logical disk equipment 200 concerned, and random access. This access pattern information 502 has many sequential accesses, and rearranging to physical disk equipment 105 with the more high sequential engine performance uses it as an index which asks for desirable logical disk equipment 200.

[0033] Next, actuation of a memory control unit 104 is explained. Drawing 6 is drawing which

expressed actuation of a memory control unit 104 to the detail. First, the actuation at the time of read/write processing is explained. In case a director 106 usually performs read/write processing, he receives the directions 600 from CPU from CPU101 via a channel 103. The directions 600 from this CPU include the assignment information 1 which specifies the logical disk equipment 200 with which the record for a lead (or light) is memorized, and the assignment information 2 which specifies the location (a truck, a sector, record) in the logical disk equipment 200 with which the record for a lead (or light) is memorized. A director 106 is access location calculation processing (610) on physical disk equipment, and computes the access location on physical disk equipment 105 using the directions 600 from said CPU, and the information 300 corresponding to logic physics. This physical disk drive-access location calculation processing (610) is explained in full detail with reference to drawing 8 later. Then, for example by lead processing, the data of the data storage location 202 on the computed physical disk equipment 105 are read out on cache memory 107, it considers as data 201, and the read-out data 201 is transmitted to a primary storage 102 through a channel 103.

[0034] Next, extraction processing of access information 500 is explained. At the time of access of the read/write processing from CPU101, a director 106 updates the access information 500 of the logical disk equipment 200 for access. Extraction of the access frequency information 501 counts up the internal counter at every access, and judges access frequency from said internal UNTA at the time of access after access progress of fixed time amount or the count of fixed. Extraction of the access pattern information 502 counts up the count of a sequential access at the internal counter at every access, and judges an access pattern from said internal counter at the time of access after access progress of fixed time amount or the count of fixed.

[0035] Next, the relocation directions 620 are explained. A customer engineer examines the need for relocation of each logical disk equipment 200 with reference to the access information 500 shown through SVP111. If there is logical disk equipment 200 which opted for relocation as a result of this examination, the relocation directions 620 will be issued to a memory control unit 104 through SVP111. These relocation directions 620 consist of directions information 1-2 which specifies two logical disk equipments 200 for relocation. The contents of examination which a customer engineer performs are the same as that of logical disk equipment relocation necessity decision processing (910) in which it explains with reference to drawing 10 with the 3rd operation gestalt mentioned later.

[0036] Next, logical disk equipment relocation processing (630) is explained. A director 106 performs logical disk equipment relocation processing (630) between two specified logical disk equipments 200 in response to said relocation directions 620. Drawing 7 is the processing flow Fig. of the logical disk equipment relocation processing section 630. At step 700, the logical disk condition 401 of two logical disk equipments 200 of having been specified of the logical disk information 400 is set up "during relocation." At step 701, the completion pointer 402 of relocation of two logical disk equipments 200 with which it was specified of the logical disk information 400 is initialized in the head location of each logical disk equipment 200. At step 702, the completion pointer 402 of relocation of two logical disk equipments 200 with which it was specified of the logical disk information 400 is checked, and if relocation of all fields is not completed, and it progressed to step 703 and has completed, it will progress to step 707.

[0037] At step 703, data transfer from physical disk equipment 105 to a cache memory 107 top is performed to the data for 1 time of the batch of relocation processing from the data location which the completion pointer 402 of relocation shows. Here, the amount of data for 1 time of a batch is determined as the least common multiple of each amount of data corresponding to one redundancy data of two logical disk equipments 200 for relocation. For example, if it rearranges between the logical disk equipment 200 of RAID5, and the logical disk equipment 200 of RAID1, since the amount of data corresponding to one redundancy data of the logical disk equipment 200 of RAID1 is "1", the amount of data for 1 time of a batch will be determined as the amount of data corresponding to one redundancy data of the logical disk equipment 200 of RAID5, i.e., the amount of data corresponding to one parity.

[0038] At step 704, when the relocation place logical disk equipment 200 of each logical disk equipment 200 for relocation is the thing of RAID level which has parity, parity is generated to

the data 201 for 1 time of the batch for [on cache memory 107] relocation. At step 705, the parity created at the data 201 and said step 704 for 1 time of a batch for relocation is written in the physical disk equipment 105 of a relocation place. [on cache memory 107] At step 706, the completion pointer 402 of relocation is carried forward by 1 time of the batch. And it returns to said step 702.

[0039] In addition, in the above-mentioned step 703,704, it transmits and doubles also to nonvolatile memory 109, and data and parity prevent data missing by the cache failure. This reason at the time of the writing in the above-mentioned step 705 For example, the inside of the data of the 1st logical disk equipment 200 and the 2nd logical disk equipment 200, Supposing the data on cache memory 107 become access impossible according to a failure in the phase which wrote the data of the 1st logical disk equipment 200 in physical disk equipment 105 (physical disk equipment 105 with which the dimension is arranged at the 2nd logical disk equipment 200) It is because the data of the 2nd logical disk equipment 200 which writing has not ended disappear (the data of the 1st logical disk equipment 200 will be overwritten as mentioned above by the physical disk equipment 105 with which the dimension is arranged at the 2nd logical disk equipment 200).

[0040] At step 707, the information 300 corresponding to logic physics is updated. That is, the logical disk configuration information 310 and the physical disk configuration information 321 are changed. At step 708, the logical disk condition 401 of the logical disk information 400 is returned to the original condition, and relocation processing (630) is ended.

[0041] Next, physical disk drive-access location calculation processing (610) is explained. Drawing 8 is the processing flow Fig. of the physical disk drive-access location calculation processing section 610. if it confirms whether the logical disk condition 401 of the logical disk equipment 200 for access of the logical disk information 400 "is rearranging" and becomes "during relocation" about it at step 800 — step 801 -- progressing -- "-- under relocation -- it is — " — if there is nothing, it will progress to step 803.

[0042] The completion pointer 402 of relocation and the access data location of the logical disk equipment 200 for access of the logical disk information 400 are compared, if an access data location becomes after the location which the completion pointer 402 of relocation points out, it will progress to step 802, and if an access data location becomes a front [location / which the completion pointer 402 of relocation points out], it will progress to step 803 at step 801.

[0043] At step 802, the logical disk equipment 200 of the relocation place of the logical disk equipment 200 concerned is made applicable to access. And it progresses to step 804.

[0044] At step 803, the logical disk equipment 200 concerned is made applicable to access.

[0045] At step 804, the access location on the physical disk equipment 105 corresponding to the logical disk equipment 200 for access is computed using the information 300 corresponding to logic physics.

[0046] According to the information processing system 1 and the memory control unit 104 concerning the above operation gestalt [1st], logical disk equipment with high access frequency is rearrangeable to more nearly high-speed physical disk equipment with the decision of a customer engineer based on access information 500. Moreover, logical disk equipment with the high ratio of a sequential access is rearrangeable to physical disk equipment with the more high sequential access engine performance. Therefore, the access engine performance can be improved.

[0047] - The operation gestalt of the 2nd operation gestalt-above 1st is transformed, and access information 500 is shown to a data processor 100 from a memory control unit 104, and a data processor 100 determines relocation necessity and you may make it take out relocation directions (about [620]) to a memory control unit 104.

[0048] - 3rd operation gestalt - The 3rd operation gestalt receives relocation directions neither from SVP111 nor a data processor 100, but a memory control unit 104 makes a self-decision.

[0049] Drawing 9 is drawing which expressed actuation of a memory control unit 104 to the detail. The difference from the 1st operation gestalt (drawing 6) is that the logical disk relocation necessity decision processing section 910 issues the relocation directions 620.

[0050] Drawing 10 is the processing flow Fig. of the above-mentioned logical disk relocation necessity decision processing section 910. A director 106 performs this logical disk relocation necessity decision processing (910) by inspecting the access information 500 of each logical disk equipment 200 a fixed period. If there is [whether there is logical disk equipment / low speed / the physical disk equipment 105 which access frequency exceeds default value and is arranged at step 1000 with reference to the access frequency information 501 on access information 500 / equipment / comparatively / (this is hereafter called 1st candidate logical disk equipment) 200, and] logical disk equipment 200 which checks and corresponds, it will progress to step 1001, and if there is nothing, it will progress to step 1005.

[0051] At step 1001, it confirms whether the ratio of a sequential access is beyond default value, with reference to the access pattern information 502 on said 1st candidate logical disk equipment 200, if it is not beyond default value, it will progress to step 1002, and with default value [beyond], it progresses to step 1004.

[0052] At step 1002, with reference to the access frequency information 501 on the logical disk equipment 200 arranged at physical disk equipment 105 more nearly high-speed than said 1st candidate logical disk equipment 200, if check and it is [whether access frequency has logical disk equipment 200 below default value (this is hereafter called 2nd candidate logical disk equipment), and], it will progress to step 1003, and if there is nothing, it will progress to step 1005.

[0053] It determines that relocation processing (630) is required between said 1st candidate logical disk equipment 200 and said 2nd candidate logical disk equipment 200, and the relocation directions 620 are taken out with step 1003. And processing is ended.

[0054] At step 1004, with reference to the access pattern information 502 on the logical disk equipment 200 arranged at physical disk equipment 105 with the sequential engine performance higher than said 1st candidate logical disk equipment 200, if check and it is [whether the ratio of a sequential access has logical disk equipment 200 below default value (this is hereafter called 2nd candidate logical disk equipment), and], it will progress to said step 1003, and if there is nothing, it will progress to said step 1002.

[0055] At step 1005, it is determined that relocation processing (630) of logical disk equipment 200 is unnecessary. And processing is ended.

[0056] According to the information processing system 1 and the memory control unit 104 concerning the above operation gestalt [3rd], based on access information 500, logical disk equipment with high access frequency is rearrangeable to more nearly high-speed physical disk equipment automatically. Moreover, logical disk equipment with the high ratio of a sequential access is rearrangeable to physical disk equipment with the more high sequential access engine performance. Therefore, the access engine performance can be improved.

[0057] - 4th operation gestalt-above-mentioned the 1- the 3rd operation gestalt -- deforming -- access information 500 — replacing with -- or — in addition, the dependability required of logical disk equipment 200 may be used for the index of relocation processing necessity decision. If dependability is used for an index, the dependability of the data on logical disk equipment 200 can be raised.

[0058]

[Effect of the Invention] According to the memory control unit of this invention, with the case of a sequential access, or a random accelerator, even when a hit ratio is low, the access engine performance can be improved. Moreover, according to the memory control unit of this invention, the dependability of data can be improved.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] It is the block diagram of the information processing system containing the memory control unit concerning the 1st operation gestalt of this invention.
[Drawing 2] It is the explanatory view of the correspondence relation between logical disk equipment and physical disk equipment.
[Drawing 3] It is the configuration instantiation Fig. of the information corresponding to logic physics.
[Drawing 4] It is the configuration instantiation Fig. of logical disk information.
[Drawing 5] It is the configuration instantiation Fig. of access information.
[Drawing 6] It is the block diagram showing actuation of the memory control unit in the 1st operation gestalt of this invention.
[Drawing 7] It is the processing flow Fig. of the logical disk equipment relocation processing section.
[Drawing 8] It is the processing flow Fig. of the physical disk drive-access location calculation processing section.
[Drawing 9] It is the block diagram showing actuation of the memory control unit in the 3rd operation gestalt of this invention.
[Drawing 10] It is the processing flow Fig. of the logical disk equipment relocation necessity decision processing section.
[Description of Notations]
1 -- Information Processing System
100 -- Data processor
101 -- CPU
102 -- Primary storage
103 -- Channel
104 -- Memory control unit
105 -- Physical disk equipment
106 -- Director
107 -- Cache memory
108 -- Cache directory
109 -- Nonvolatile memory
110 -- Nonvolatile memory management information
111 -- SVP
200 -- Logical disk equipment
201 -- Data
202 -- Data storage location
300 -- Information corresponding to logic physics
400 -- Logical disk information
500 -- Access information
600 -- Directions from CPU
610 -- The access location calculation processing section on physical disk equipment

620 -- Directions information
630 -- Logical disk equipment relocation processing section
910 -- Logical disk relocation necessity decision processing section

[Translation done.]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2000－293317
（Ｐ2000－293317Ａ）
(43)公開日　平成12年10月20日(2000.10.20)

(51)Int.Cl.'		識別記号	Ｆ Ｉ		テーマコード（参考）
Ｇ０６Ｆ	3/06	３０２	Ｇ０６Ｆ	3/06	３０２Ｊ
		３０１			３０１Ｓ
		５４０			５４０
	12/16	３２０		12/16	３２０Ｌ
Ｇ１１Ｂ	19/02	５０１	Ｇ１１Ｂ	19/02	５０１Ｋ

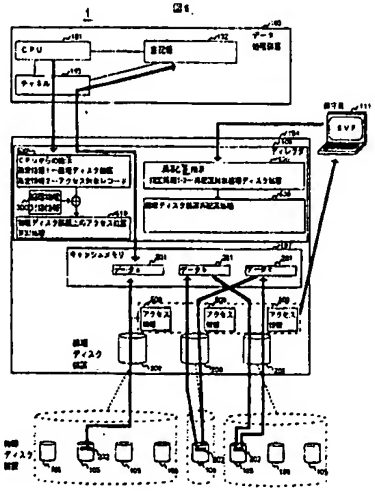
審査請求　未請求　請求項の数10　ＯＬ　（全 14 頁）　最終頁に続く

(21)出願番号　特願2000－66061(Ｐ2000－66061)
(62)分割の表示　特願平8－85370の分割
(22)出願日　平成８年４月８日(1996.4.8)

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(72)発明者　山本　康友
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内
(72)発明者　山本　彰
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内
(72)発明者　佐藤　幸夫
神奈川県小田原市国府津2880番地　株式会社日立製作所ストレージシステム事業部内
(74)代理人　100095511
弁理士　有近　紳志郎

(54)【発明の名称】　記憶制御装置

(57)【要約】
【課題】　アクセス性能を向上する。
【解決手段】　各論理ディスク装置２００に対するアクセス情報５００を採取し、そのアクセス情報５００をＳＶＰ１１１を通じて保守員に提示する。保守員の再配置指示６２０があると、指示された２つの論理ディスク装置２００の間で物理ディスク装置１０５を配置替えし、全データを連続的に格納し直す。
【効果】　アクセス頻度の高い論理ディスク装置をより高速な物理ディスク装置へ再配置することが出来る。シーケンシャルアクセスの比率の高い論理ディスク装置をよりシーケンシャルアクセス性能の高い物理ディスク装置へ再配置することが出来る。



【特許請求の範囲】
【請求項１】　データ処理装置が直接アクセスを行う論理的記憶装置を実際にデータを記憶する物理的記憶装置に配置し、前記データ処理装置と前記物理的記憶装置の間のデータ転送を制御する記憶制御装置において、
予め定めた指標に基づいて前記論理的記憶装置を前記物理的記憶装置に再配置すると共に再配置先の物理的記憶装置にデータを連続的に格納する論理的記憶装置再配置手段を有することを特徴とする記憶制御装置。
【請求項２】　データ処理装置が直接アクセスを行う論理的記憶装置と実際にデータを記憶する物理的記憶装置とを対応付け、前記データ処理装置と前記物理的記憶装置の間のデータ転送を制御する記憶制御装置において、前記データ転送の制御の運用中にデータ処理装置の論理的記憶装置へのアクセス情報を指標として採取するアクセス情報採取手段と、前記指標に基づいて前記論理的記憶装置を前記物理的記憶装置に再配置すると共に再配置先の物理的記憶装置にデータを連続的に格納する論理的記憶装置再配置手段とを有することを特徴とする記憶制御装置。
【請求項３】　請求項２に記載の記憶制御装置において、前記アクセス情報が、前記データ処理装置から前記論理的記憶装置へのアクセス頻度情報を含むことを特徴とする記憶制御装置。
【請求項４】　請求項２または請求項３に記載の記憶制御装置において、前記アクセス情報が、前記データ処理装置から前記論理的記憶装置へのアクセスパターン情報を含むことを特徴とする記憶制御装置。
【請求項５】　請求項１に記載の記憶制御装置において、前記指標が、前記論理的記憶装置に求められる信頼性であることを特徴とする記憶制御装置。
【請求項６】　請求項１から請求項５のいずれかに記載の記憶制御装置において、前記指標を保守員に提示する指標提示手段と、保守員からの再配置指示を受け付ける再配置指示受付手段とを具備したことを特徴とする記憶制御装置。
【請求項７】　請求項１から請求項５のいずれかに記載の記憶制御装置において、データ処理装置からの再配置指示を受け付ける再配置指示受付手段を具備したことを特徴とする記憶制御装置。
【請求項８】　請求項１から請求項５のいずれかに記載の記憶制御装置において、前記指標に基づいて再配置の要否を決定する再配置要否決定手段を具備したことを特徴とする記憶制御装置。
【請求項９】　請求項１から請求項８のいずれかに記載の記憶制御装置において、再配置中の論理的記憶装置にデータ処理装置からのアクセスがあったとき、再配置中の論理的記憶装置の再配置完了領域と再配置未完領域とを識別し、前記アクセス位置が前記再配置完了領域ならば再配置先の論理的記憶装置にアクセスさせ、前記アクセス位置が前記再配置未完領域ならば当該論理的記憶装置にアクセスさせるアクセス位置切替手段をさらに具備したことを特徴とする記憶制御装置。
【請求項１０】　データ処理装置がデータアクセスのために認識する論理的な記憶装置のデータを保持するための複数の物理的な記憶装置と、前記複数の物理装置とデータ処理装置との間のデータ転送を制御する記憶制御装置とを有する記憶装置システムにおいて、
前記記憶制御装置は、前記データ処理装置によるアクセス状況を取得し、前記アクセス状況に基づいて前記論理的な記憶装置のデータを第１の物理的な記憶装置から第２の物理的な記憶装置に移動させることを特徴とする記憶装置システム。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、記憶制御装置に関し、さらに詳しくは、シーケンシャルアクセスの場合やランダムアクセルでヒット率が低い場合でもアクセス性能を向上することが出来る記憶制御装置およびデータの信頼性を向上することが出来る記憶制御装置に関する。
特に、本発明は、ディスクアレイ向きの高機能ディスク装置、その高機能ディスク装置とディスク制御装置とにより構成される記憶装置サブシステム、およびその記憶装置サブシステムとデータ処理装置とにより構成される情報処理システムに有用である。
【０００２】
【従来の技術】シカゴのイリノイ大学で開かれた「ACM SIGMOD」会議において発表された論文「D.Patterson,C.gibson,and R.H.Kartz;A Case for Redundant Arrays of Inexpensive Disks (RAID),ACM SIGMOD Conference,Chicago,IL,(June 1988),pp.109-116」は、ディスクアレイ上のデータ配置に関する技術を開示している。
【０００３】また、特開平７－８４７３２号公報では、ディスク装置の一部をディスクキャッシュの如く用いる技術が開示されている。具体的には、ディスク装置を一時的にデータを格納するテンポラリ領域と最終的にデータを書き込む領域とに分け、更新データはパリティを生成せずに一旦テンポラリ領域に二重書きし、非同期にパリティ生成し、最終領域に書き込む。
【０００４】一方、電気情報通信学会技術研究報告「ＤＥ９５－６８（茂木他：Hot Mirroring を用いたディスクアレイのディスク故障時の性能評価、１９９５年１２月、電気情報通信学会技報 Vol.95-No.407、pp.19-24）」には、アクセス頻度の違いにより、データを保持するＲＡＩＤレベルを動的に変更する技術が開示されている。具体的には、ディスク装置をＲＡＩＤ１構成の部分とＲＡＩＤ５構成の部分に分け、ライトアクセスのあったデータを優先的にＲＡＩＤ１構成の部分に格納するようにデータの格納位置を動的に変更することにより、アクセス頻度の高いデータはＲＡＩＤ１構成の部分に格

納し、アクセス頻度の低いものはＲＡＩＤ５構成の部分に格納するように出来る。この技術によれば、記憶容量の異なる物理ディスク装置やＲＡＩＤレベルの異なる物理ディスク装置を記憶装置サブシステム内で混在させることが可能であり、論理ディスク装置内のデータを、そのアクセス頻度やアクセスパターンなどの指標に基づいて、任意の物理ディスク装置に格納することが出来る。また、アクセス頻度の高いデータを、より高速な物理ディスク装置に格納するように、動的に格納位置を変更することも出来る。なお、ＲＡＩＤ１のディスクアレイは、データ処理装置からの書き込みデータに対して、その複製をミラーと呼ばれる副ディスク装置に書き込み、データの信頼性を確保する。冗長データが元のデータの複製であるため、冗長データ作成のオーバヘッドが小さく、アクセス性能が良い。但し、物理的記憶装置の使用効率は、５０％と低い。一方、ＲＡＩＤ５のディスクアレイは、データ処理装置からの複数の書き込みデータに対して、パリティと呼ばれる冗長データを作成する。パリティ作成時に更新前データと更新前パリティのリードが必要なため、冗長データ作成のオーバヘッドが大きく、アクセス性能は悪い。但し、複数のデータに対して１つのパリティを作成するため、記憶装置の使用効率はＲＡＩＤ１に比べ高い。

【０００５】

【発明が解決しようとする課題】上記従来技術では、アクセスするデータ単位でデータの格納位置の変更を行うため、データ処理装置が直接アクセスを行う論理ディスク装置上では連続なデータが、実際にデータを記憶する物理ディスク装置上では非連続となってしまう。このため、一連のデータをリード／ライトするシーケンシャルアクセスの場合、実際には複数データをまとめてリード／ライトできなくなり、アクセス性能の低下を招く問題点がある。

【０００６】一方、上記報告「ＤＥ９５－６８」の従来技術では、ライトの度に、アクセス頻度が低いと判断したデータをＲＡＩＤ１構成の部分からＲＡＩＤ５構成の部分に移し、空いたＲＡＩＤ１構成の部分にライトデータを書き込むため、アクセスパターンがランダムアクセルでヒット率が低い場合には、ＲＡＩＤ１構成の部分に移したデータの多くは再びＲＡＩＤ５構成の部分に戻されることになる。このため、ヒット率が低い場合、アクセス性能の向上は期待できず、逆にデータを移す処理のオーバヘッドがアクセス性能の低下を引き起こす問題点がある。

【０００７】また、上記の従来技術では、データの信頼性の向上については全く考慮されていない問題点がある。

【０００８】そこで、本発明の第１の目的は、シーケンシャルアクセスの場合やランダムアクセルでヒット率が低い場合でも、アクセス性能を向上することが出来る記憶制御装置を提供することにある。また、本発明の第２の目的は、データの信頼性を向上することが出来る記憶制御装置を提供することにある。

【０００９】

【課題を解決するための手段】第１の観点では、本発明は、データ処理装置が直接アクセスを行う論理的記憶装置を実際にデータを記憶する物理的記憶装置に配置し、前記データ処理装置と前記物理的記憶装置の間のデータ転送を制御する記憶制御装置において、予め定めた指標に基づいて前記論理的記憶装置を前記物理的記憶装置に再配置すると共に再配置先の物理的記憶装置にデータを連続的に格納する論理的記憶装置再配置手段を有することを特徴とする記憶制御装置を提供する。上記第１の観点による記憶制御装置では、アクセスするデータ単位でデータの格納位置の変更を行うのではなく、論理的記憶装置を単位として物理的記憶装置への再配置を行い、且つ、再配置先の物理的記憶装置にデータを連続的に格納する。従って、シーケンシャルアクセスの場合でも、アクセス性能を向上することが出来る。また、ライトの度にデータの格納位置の変更を行うのではなく、予め定めた指標に基づいて前記再配置を行うから、ランダムアクセルでヒット率が低い場合でも、アクセス性能を向上することが出来る。

【００１０】第２の観点では、本発明は、データ処理装置が直接アクセスを行う論理的記憶装置と実際にデータを記憶する物理的記憶装置とを対応付け、前記データ処理装置と前記物理的記憶装置の間のデータ転送を制御する記憶制御装置において、前記データ転送の制御の運用中にデータ処理装置の論理的記憶装置へのアクセス情報を指標として採取するアクセス情報採取手段と、前記指標に基づいて前記論理的記憶装置を前記物理的記憶装置に再配置すると共に再配置先の物理的記憶装置にデータを連続的に格納する論理的記憶装置再配置手段とを有することを特徴とする記憶制御装置を提供する。上記第２の観点による記憶制御装置では、アクセスするデータ単位でデータの格納位置の変更を行うのではなく、論理的記憶装置を単位として物理的記憶装置への再配置を行い、且つ、再配置先の物理的記憶装置にデータを連続的に格納する。従って、シーケンシャルアクセスの場合でも、アクセス性能を向上することが出来る。また、ライトの度にデータの格納位置の変更を行うのではなく、アクセス情報を採取し、それを統計的に利用して前記再配置を行うから、ランダムアクセルでヒット率が低い場合でも、アクセス性能を向上することが出来る。

【００１１】第３の観点では、本発明は、上記構成の記憶制御装置において、前記アクセス情報が、前記データ処理装置から前記論理的記憶装置へのアクセス頻度情報を含むことを特徴とする記憶制御装置を提供する。上記第３の観点による記憶制御装置では、アクセス頻度の高い論理的記憶装置をより高速な物理的記憶装置へ再配置することが出来る。従って、アクセス性能を向上することが出来る。

【００１２】第４の観点では、本発明は、上記構成の記憶制御装置において、前記アクセス情報が、前記データ処理装置から前記論理的記憶装置へのアクセスパターン情報を含むことを特徴とする記憶制御装置を提供する。上記第４の観点による記憶制御装置では、シーケンシャルアクセスの比率の高い論理的記憶装置をよりシーケンシャルアクセス性能の高い物理的記憶装置へ再配置することが出来る。従って、アクセス性能を向上することが出来る。

【００１３】第５の観点では、本発明は、上記構成の記憶制御装置において、前記指標が、前記論理的記憶装置に求められる信頼性であることを特徴とする記憶制御装置を提供する。上記第５の観点による記憶制御装置では、信頼性が高いことが求められる論理的記憶装置をより信頼性の高い物理的記憶装置へ再配置することが出来る。従って、データの信頼性を向上することが出来る。

【００１４】第６の観点では、本発明は、上記構成の記憶制御装置において、前記指標を保守員に提示する指標提示手段と、保守員からの再配置指示を受け付ける再配置指示受付手段とを具備したことを特徴とする記憶制御装置を提供する。上記第６の観点による記憶制御装置では、保守員が再配置指示を入力できるため、非常に柔軟に前記再配置を行うことが出来る。

【００１５】第７の観点では、本発明は、上記構成の記憶制御装置において、データ処理装置からの再配置指示を受け付ける再配置指示受付手段を具備したことを特徴とする記憶制御装置を提供する。上記第７の観点による記憶制御装置では、データ処理装置が再配置指示を入力できるため、保守員では判断不可能な高度の条件下で前記再配置を行うことが出来る。

【００１６】第８の観点では、本発明は、上記構成の記憶制御装置において、前記指標に基づいて再配置の要否を決定する再配置要否決定手段を具備したことを特徴とする記憶制御装置を提供する。上記第８の観点による記憶制御装置では、記憶制御装置が再配置指示を自己決定するため、保守員やデータ処理装置に負担をかけなくて済む。

【００１７】第９の観点では、本発明は、上記構成の記憶制御装置において、再配置中の論理的記憶装置にデータ処理装置からのアクセスがあったとき、再配置中の論理的記憶装置の再配置完了領域と再配置未完領域とを識別し、前記アクセス位置が前記再配置完了領域ならば再配置先の論理的記憶装置にアクセスさせ、前記アクセス位置が前記再配置未完領域ならば当該論理的記憶装置にアクセスさせるアクセス位置切替手段を具備したことを特徴とする記憶制御装置を提供する。上記第９の観点による記憶制御装置では、再配置中の論理的記憶装置の再配置完了領域と再配置未完領域とを識別し、データ処理装置からのアクセス位置を切り替えるから、データ処理装置と物理的記憶装置の間のデータ転送を運用中に再配置を行うことが出来る。

【００１８】第１０の観点では、本発明は、データ処理装置がデータアクセスのために認識する論理的な記憶装置のデータを保持するための複数の物理的な記憶装置と、前記複数の物理装置とデータ処理装置との間のデータ転送を制御する記憶制御装置とを有する記憶装置システムにおいて、前記記憶制御装置は、前記データ処理装置によるアクセス状況を取得し、前記アクセス状況に基づいて前記論理的な記憶装置のデータを第１の物理的な記憶装置から第２の物理的な記憶装置に移動させることを特徴とする記憶装置システムを提供する。上記第１０の観点による記憶制御装置では、データ処理装置によるアクセス状況に応じて論理的記憶装置のデータを第１から第２の物理的な記憶装置へと移動させるから、アクセス性能を向上することが出来る。

【００１９】

【発明の実施の形態】以下、本発明の実施形態を説明する。なお、これにより本発明が限定されるものではない。

【００２０】－第１の実施形態－

第１の実施形態は、各論理ディスク装置のアクセス情報を記憶制御装置で採取し、ＳＶＰ（サービスプロセッサ）を通じて保守員に提示し、このアクセス情報に基づく保守員の再配置指示により、論理ディスク装置の物理ディスク装置への再配置を行うものである。

【００２１】図１は、本発明の第１の実施形態にかかる記憶制御装置を含む情報処理システムのブロック図である。この情報処理システム１は、データ処理装置１００と、記憶制御装置１０４と、１台以上の物理ディスク装置１０５と、ＳＶＰ１１１とを接続してなっている。

【００２２】前記データ処理装置１００は、ＣＰＵ１０１と、主記憶１０２と、チャネル１０３とを有している。

【００２３】前記記憶制御装置１０４は、１つ以上のディレクタ１０６と、キャッシュメモリ１０７と、ディレクトリ１０８と、不揮発性メモリ１０９と、不揮発性メモリ管理情報１１０と、論理物理対応情報３００と、論理ディスク装置情報４００と、アクセス情報５００を有している。前記ディレクタ１０６は、データ処理装置１００のチャネル１０３と物理ディスク装置１０５の間のデータ転送、データ処理装置１００のチャネル１０３と前記キャッシュメモリ１０７の間のデータ転送および前記キャッシュメモリ１０７と物理ディスク装置１０５の間のデータ転送を行う。前記キャッシュメモリ１０７には、物理ディスク装置１０５の中のアクセス頻度の高いデータをロードしておく。このロード処理は、前記ディレクタ１０６が実行する。ロードするデータの具体例は、データ処理装置１００のＣＰＵ１０１のアクセス対

象データや，このアクセス対象データと物理ディスク装置１０５上の格納位置が近いデータ等である。前記ディレクトリ１０８は、前記キャッシュメモリ１０７の管理情報を格納する。前記不揮発性メモリ１０９は、前記キャッシュメモリ１０７と同様に、物理ディスク装置１０５の中のアクセス頻度の高いデータをロードしておく。前記不揮発性メモリ管理情報１１０は、前記不揮発性メモリ１０９の管理情報を格納する。前記論理物理対応情報３００は、各論理ディスク装置（図2の２００）が配置されている物理ディスク装置１０５上の位置および各物理ディスク装置１０５に配置されている論理ディスク装置（図2の２００）を示す情報である。この情報を用いて、データ処理装置１００のＣＰＵ１０１のアクセス対象データの物理ディスク装置１０５上の格納領域の算出などを行う。前記論理ディスク装置情報４００は、各論理ディスク装置（図2の２００）のアクセス可否等の状態を示す。前記アクセス情報５００は、各論理ディスク装置（図2の２００）のアクセス頻度やアクセスパターンなどの情報である。

【００２４】論理物理対応情報３００と論理ディスク情報４００は、電源断などによる消失を防ぐために不揮発の媒体に記録する。

【００２５】前記物理ディスク装置１０５は、データを記録する媒体と、記録されたデータを読み書きする装置とから構成される。

【００２６】前記ＳＶＰ１１１は、アクセス情報５００の保守員への提示や，保守員からの再配置指示６２０の入力の受け付けを行う。また、保守員からの情報処理システム１への指示の発信や，情報処理システム１の障害状態等の保守員への提示を行う。

【００２７】図2は、論理ディスク装置２００と物理ディスク装置１０５の関連を表わした図である。論理ディスク装置２００は、データ処理装置１００のＣＰＵ１０１が直接アクセスする見掛け上のディスク装置で、アクセス対象データが実際に格納される物理ディスク装置１０５と対応している。論理ディスク装置２００上のデータは、シーケンシャルアクセスを考慮して、物理ディスク装置１０５上に連続的に配置されている。論理ディスク装置２００のデータが配置されている物理ディスク装置１０５がディスクアレイ構成の場合、該論理ディスク装置２００は複数の物理ディスク装置１０５と対応する。また、物理ディスク装置１０５の容量が論理ディスク装置２００より大きく、複数の論理ディスク装置のデータを１台の物理ディスク装置１０５に格納できる場合には、該物理ディスク装置１０５は複数の論理ディスク装置２００と対応する。この論理ディスク装置２００と物理ディスク装置１０５の対応は前記論理物理対応情報３００で管理される。例えば、データ処理装置１００のＣＰＵ１０１が論理ディスク装置２００のデータ２０１をリードする時、記憶制御装置１０４で論理物理対応情報３００に基づき論理ディスク装置２００に対応する物理ディスク装置１０５を求め、その物理ディスク装置１０５の領域内のデータ格納位置２０２を求め、データ転送を行う。

【００２８】図3は、論理物理対応情報３００を表わした図である。論理物理対応情報３００は、論理ディスク構成情報３１０と、物理ディスク構成情報３２０とから構成される。前記論理ディスク構成情報３１０は、各論理ディスク装置２００が配置されている物理ディスク装置１０５上の領域に関する情報であり、論理ディスク装置２００から対応する物理ディスク装置１０５を求める時に用いる。一方、前記物理ディスク構成情報３２０は、各物理ディスク装置１０５に配置されている論理ディスク装置２００に関する情報で、物理ディスク装置１０５から対応する論理ディスク装置２００を求める時に用いる。

【００２９】前記論理ディスク構成情報３１０は、物理ディスク装置グループ３１１，ＲＡＩＤ構成３１２および開始位置３１３の組を、論理ディスク装置２００の数だけ有している。前記物理ディスク装置グループ３１１は、当該論理ディスク装置２００が配置されている物理ディスク装置１０５を示す情報である。前記ＲＡＩＤ構成３１２は、前記物理ディスク装置グループ３１１のＲＡＩＤレベルを示す。前記開始位置３１３は、当該論理ディスク装置２００が物理ディスク装置１０５上で配置されている先頭位置を示す。

【００３０】前記物理ディスク構成情報３２０は、論理ディスク装置グループ３２１を、物理ディスク装置１０５の数だけ有している。前記論理ディスク装置グループ３２１は、当該物理ディスク装置１０５に配置されている論理ディスク装置２００を示す。

【００３１】図4は、論理ディスク情報４００を表わした図である。論理ディスク情報４００は、論理ディスク状態４０１と再配置完了ポインタ４０２とを、論理ディスク装置２００の数だけ有している。前記論理ディスク状態４０１は、「正常」「閉塞」「フォーマット中」「再配置中」などの論理ディスク装置２００の状態を表わす。前記再配置完了ポインタ４０２は、前記論理ディスク状態４０１が「再配置中」の時のみ有効な情報で、当該論理ディスク装置２００の再配置処理を完了している領域の次の位置すなわち当該論理ディスク装置２００が未だ再配置処理を終えていない領域の先頭位置を示す。「再配置中」におけるデータアクセス時、再配置完了ポインタ４０２よりも前の領域へのアクセスの場合には、再配置後の物理ディスク装置１０５へアクセスしなければならない。一方、再配置完了ポインタ４０２以後の領域へのアクセスの場合には、再配置前の物理ディスク装置１０５へアクセスしなければならない。

【００３２】図5は、アクセス情報５００を表わしている。アクセス情報５００は、アクセス頻度情報５０１とアクセスパターン情報５０２とを、論理ディスク装置２００の数だけ有している。このアクセス情報５００は、記憶制御装置１０４，データ処理装置１００，ＳＶＰ１１１のいずれからも参照することが出来る。前記アクセス頻度情報５０１は、単位時間あたりの当該論理ディスク装置２００へのアクセス回数を管理する。このアクセス頻度情報５０１は、各論理ディスク装置２００の中でアクセス頻度の高いもの又は低いものを求める指標として用いる。前記アクセスパターン情報５０２は、当該論理ディスク装置２００へのシーケンシャルアクセスとランダムアクセスの割合を管理する。このアクセスパターン情報５０２は、シーケンシャルアクセスが多く、よりシーケンシャル性能の高い物理ディスク装置１０５に再配置するのが望ましい論理ディスク装置２００を求める指標として用いる。

【００３３】次に、記憶制御装置１０４の動作を説明する。図6は、記憶制御装置１０４の動作を詳細に表わした図である。まず、リード／ライト処理時の動作について説明する。ディレクタ１０６は、通常リード／ライト処理を実行する際、ＣＰＵ１０１からチャネル１０３を経由してＣＰＵからの指示６００を受け取る。このＣＰＵからの指示６００は、リード（またはライト）対象のレコードが記憶されている論理ディスク装置２００を指定する指定情報1と、リード（またはライト）対象のレコードが記憶されている論理ディスク装置２００内の位置（トラック，セクタ，レコード）を指定する指定情報2とを含んでいる。ディレクタ１０６は、物理ディスク装置上のアクセス位置算出処理（６１０）で、前記ＣＰＵからの指示６００と論理物理対応情報３００とを用いて、物理ディスク装置１０５上でのアクセス位置を算出する。この物理ディスク装置アクセス位置算出処理（６１０）については図8を参照して後で詳述する。その後、たとえばリード処理では、算出した物理ディスク装置１０５上のデータ格納位置２０２のデータをキャッシュメモリ１０７上に読み上げてデータ２０１とし、その読み上げたデータ２０１をチャネル１０３を通じて主記憶１０２に転送する。

【００３４】次に、アクセス情報５００の採取処理について説明する。ＣＰＵ１０１からのリード／ライト処理のアクセス時に、ディレクタ１０６は、アクセス対象論理ディスク装置２００のアクセス情報５００を更新する。アクセス頻度情報５０１の採取は、例えば、アクセスの度に内部カウンタをカウントアップしていき、一定時間または一定回数のアクセス経過後のアクセス時に、前記内部カウンタからアクセス頻度を判定する。アクセスパターン情報５０２の採取は、例えば、アクセスの度に内部カウンタにシーケンシャルアクセス回数をカウントアップしていき、一定時間または一定回数のアクセス経過後のアクセス時に、前記内部カウンタからアクセスパターンを判定する。

【００３５】次に、再配置指示６２０を説明する。保守員は、ＳＶＰ１１１を通じて提示されたアクセス情報５００を参照して、各論理ディスク装置２００の再配置の必要性を検討する。この検討の結果、再配置を決定した論理ディスク装置２００があれば、ＳＶＰ１１１を通じて記憶制御装置１０４に対して再配置指示６２０を出す。この再配置指示６２０は、再配置対象の論理ディスク装置２００を２つ指定する指示情報１－２からなる。保守員が行う検討の内容は、後述する第3の実施形態で図１０を参照して説明する論理ディスク装置再配置要否決定処理（９１０）と同様である。

【００３６】次に、論理ディスク装置再配置処理（６３０）を説明する。ディレクタ１０６は、前記再配置指示６２０を受けて、指定された２つの論理ディスク装置２００の間で論理ディスク装置再配置処理（６３０）を行う。図7は、論理ディスク装置再配置処理部６３０の処理フロー図である。ステップ７００では、論理ディスク情報４００のうちの指定された2つの論理ディスク装置２００の論理ディスク状態４０１を「再配置中」に設定する。ステップ７０１では、論理ディスク情報４００のうちの指定された2つの論理ディスク装置２００の再配置完了ポインタ４０２を各論理ディスク装置２００の先頭位置に初期化する。ステップ７０２では、論理ディスク情報４００のうちの指定された2つの論理ディスク装置２００の再配置完了ポインタ４０２をチェックし、全領域の再配置が完了していなければステップ７０３へ進み、完了していればステップ７０７へ進む。

【００３７】ステップ７０３では、再配置完了ポインタ４０２が示すデータ位置から再配置処理の１回の処理単位分のデータに対して物理ディスク装置１０５からキャッシュメモリ１０７上へのデータ転送を行う。ここで、１回の処理単位分のデータ量は、再配置対象の２つの論理ディスク装置２００の冗長データ１つに対応する各データ量の最小公倍数に決定される。たとえば、再配置をＲＡＩＤ５の論理ディスク装置２００とＲＡＩＤ１の論理ディスク装置２００の間で行うならば、ＲＡＩＤ１の論理ディスク装置２００の冗長データ１つに対応するデータ量は“１”であるから、１回の処理単位分のデータ量は、ＲＡＩＤ５の論理ディスク装置２００の冗長データ１つに対応するデータ量すなわちパリティ１つに対応するデータ量に決定される。

【００３８】ステップ７０４では、再配置対象の各論理ディスク装置２００の再配置先論理ディスク装置２００がパリティを有するＲＡＩＤレベルのものである場合、キャッシュメモリ１０７上の再配置対象の１回の処理単位分のデータ２０１に対してパリティを生成する。ステップ７０５では、キャッシュメモリ１０７上の再配置対象の１回の処理単位分のデータ２０１および前記ステップ７０４で作成したパリティを、再配置先の物理ディスク装置１０５へ書き込む。ステップ７０６では、１回の

処理単位分だけ再配置完了ポインタ４０２を進める。そして、前記ステップ７０２に戻る。

【００３９】なお、上記ステップ７０３，７０４において、データおよびパリティは、不揮発性メモリ１０９にも転送して二重化し、キャッシュ障害によるデータ消失を防ぐ。この理由は、上記ステップ７０５での書き込み時に、例えば、第１の論理ディスク装置２００と第２の論理ディスク装置２００のデータのうち、第１の論理ディスク装置２００のデータを物理ディスク装置１０５（元は第２の論理ディスク装置２００に配置されていた物理ディスク装置１０５）へ書き込んだ段階で障害によりキャッシュメモリ１０７上のデータがアクセス不能になったとすると、書き込みが終了してない第２の論理ディスク装置２００のデータが消失するからである（元は第２の論理ディスク装置２００に配置されていた物理ディスク装置１０５には、上記のように第１の論理ディスク装置２００のデータが上書きされてしまっている）。

【００４０】ステップ７０７では、論理物理対応情報３００を更新する。すなわち、論理ディスク構成情報３１０と物理ディスク構成情報３２１を変更する。ステップ７０８では、論理ディスク情報４００の論理ディスク状態４０１を元の状態に戻し、再配置処理（６３０）を終了する。

【００４１】次に、物理ディスク装置アクセス位置算出処理（６１０）を説明する。図８は、物理ディスク装置アクセス位置算出処理部６１０の処理フロー図である。ステップ８００では、論理ディスク情報４００のうちのアクセス対象論理ディスク装置２００の論理ディスク状態４０１が「再配置中」であるか否かをチェックし、「再配置中」ならばステップ８０１に進み、「再配置中で」なければステップ８０３に進む。

【００４２】ステップ８０１では、論理ディスク情報４００のうちのアクセス対象論理ディスク装置２００の再配置完了ポインタ４０２とアクセスデータ位置とを比較し、アクセスデータ位置が再配置完了ポインタ４０２の指す位置以後ならばステップ８０２に進み、アクセスデータ位置が再配置完了ポインタ４０２の指す位置より前ならばステップ８０３に進む。

【００４３】ステップ８０２では、当該論理ディスク装置２００の再配置先の論理ディスク装置２００をアクセス対象にする。そして、ステップ８０４へ進む。

【００４４】ステップ８０３では、当該論理ディスク装置２００をアクセス対象とする。

【００４５】ステップ８０４では、アクセス対象の論理ディスク装置２００に対応した物理ディスク装置１０５上でのアクセス位置を、論理物理対応情報３００を用いて算出する。

【００４６】以上の第１の実施形態にかかる情報処理システム１および記憶制御装置１０４によれば、アクセス情報５００に基づく保守員の判断により、アクセス頻度の高い論理ディスク装置をより高速な物理ディスク装置へ再配置することが出来る。また、シーケンシャルアクセスの比率の高い論理ディスク装置をよりシーケンシャルアクセス性能の高い物理ディスク装置へ再配置することが出来る。従って、アクセス性能を向上することが出来る。

【００４７】－第２の実施形態－

上記第１の実施形態を変形して、記憶制御装置１０４からアクセス情報５００をデータ処理装置１００に提示し、データ処理装置１００が再配置要否を決定し記憶制御装置１０４に再配置指示（６２０相当）を出すようにしてもよい。

【００４８】－第３の実施形態－

第３の実施形態は、再配置指示をＳＶＰ１１１やデータ処理装置１００から受けるのではなく、記憶制御装置１０４が自己決定するものである。

【００４９】図９は、記憶制御装置１０４の動作を詳細に表わした図である。第１の実施形態（図６）との違いは、論理ディスク再配置要否決定処理部９１０が再配置指示６２０を出すことである。

【００５０】図１０は、上記論理ディスク再配置要否決定処理部９１０の処理フロー図である。この論理ディスク再配置要否決定処理（９１０）は、ディレクタ１０６が一定周期で各論理ディスク装置２００のアクセス情報５００を検査して行う。ステップ１０００では、アクセス情報５００のアクセス頻度情報５０１を参照し、アクセス頻度が規定値を超え且つ配置されている物理ディスク装置１０５が比較的低速なものである論理ディスク装置（以下、これを第１候補論理ディスク装置という）２００があるか否かをチェックし、該当する論理ディスク装置２００があればステップ１００１へ進み、なければステップ１００５へ進む。

【００５１】ステップ１００１では、前記第１候補論理ディスク装置２００のアクセスパターン情報５０２を参照し、シーケンシャルアクセスの比率が規定値以上であるか否かをチェックし、規定値以上でなければステップ１００２へ進み、規定値以上であればステップ１００４へ進む。

【００５２】ステップ１００２では、前記第１候補論理ディスク装置２００より高速な物理ディスク装置１０５に配置されている論理ディスク装置２００のアクセス頻度情報５０１を参照し、アクセス頻度が規定値以下の論理ディスク装置（以下、これを第２候補論理ディスク装置という）２００があるか否かをチェックし、あればステップ１００３へ進み、なければステップ１００５へ進む。

【００５３】ステップ１００３では、前記第１候補論理ディスク装置２００と前記第２候補論理ディスク装置２００の間で再配置処理（６３０）が必要であると決定し、再配置指示６２０を出す。そして、処理を終了する。

【００５４】ステップ１００４では、前記第１候補論理ディスク装置２００よりシーケンシャル性能の高い物理ディスク装置１０５に配置されている論理ディスク装置２００のアクセスパターン情報５０２を参照し、シーケンシャルアクセスの比率が規定値以下の論理ディスク装置（以下、これを第２候補論理ディスク装置という）２００があるか否かをチェックし、あれば前記ステップ１００３へ進み、なければ前記ステップ１００２へ進む。

【００５５】ステップ１００５では、論理ディスク装置２００の再配置処理（６３０）は不要であると決定する。そして、処理を終了する。

【００５６】以上の第３の実施形態にかかる情報処理システム１および記憶制御装置１０４によれば、アクセス情報５００に基づいて自動的に、アクセス頻度の高い論理ディスク装置をより高速な物理ディスク装置へ再配置することが出来る。また、シーケンシャルアクセスの比率の高い論理ディスク装置をよりシーケンシャルアクセス性能の高い物理ディスク装置へ再配置することが出来る。従って、アクセス性能を向上することが出来る。

【００５７】－第４の実施形態－

上記第１～第３の実施形態を変形して、アクセス情報５００に代えて又は加えて、論理ディスク装置２００に要求される信頼性を再配置処理要否決定の指標に用いてもよい。信頼性を指標に用いれば、論理ディスク装置２００上のデータの信頼性を向上させることが出来る。

【００５８】

【発明の効果】本発明の記憶制御装置によれば、シーケンシャルアクセスの場合やランダムアクセルでヒット率が低い場合でも、アクセス性能を向上することが出来る。また、本発明の記憶制御装置によれば、データの信頼性を向上することが出来る。

【図面の簡単な説明】

【図１】本発明の第１の実施形態にかかる記憶制御装置を含む情報処理システムのブロック図である。

【図２】論理ディスク装置と物理ディスク装置との対応関係の説明図である。

【図３】論理物理対応情報の構成例示図である。

【図４】論理ディスク情報の構成例示図である。

【図５】アクセス情報の構成例示図である。

【図６】本発明の第１の実施形態における記憶制御装置の動作を示すブロック図である。

【図７】論理ディスク装置再配置処理部の処理フロー図である。

【図８】物理ディスク装置アクセス位置算出処理部の処理フロー図である。

【図９】本発明の第３の実施形態における記憶制御装置の動作を示すブロック図である。

【図１０】論理ディスク装置再配置要否決定処理部の処理フロー図である。

【符号の説明】

１　…情報処理システム
１００…データ処理装置
１０１…ＣＰＵ
１０２…主記憶
１０３…チャネル
１０４…記憶制御装置
１０５…物理ディスク装置
１０６…ディレクタ
１０７…キャッシュメモリ
１０８…キャッシュディレクトリ
１０９…不揮発性メモリ
１１０…不揮発性メモリ管理情報
１１１…ＳＶＰ
２００…論理ディスク装置
２０１…データ
２０２…データ格納位置
３００…論理物理対応情報
４００…論理ディスク情報
５００…アクセス情報
６００…ＣＰＵからの指示
６１０…物理ディスク装置上のアクセス位置算出処理部
６２０…指示情報
６３０…論理ディスク装置再配置処理部
９１０…論理ディスク再配置要否決定処理部

【図４】

図４

論理ディスク情報

400

論理ディスク装置の数だけ用意

論理ディスク状態 401
再配置完了ポインタ 402

【図５】

図５

アクセス情報

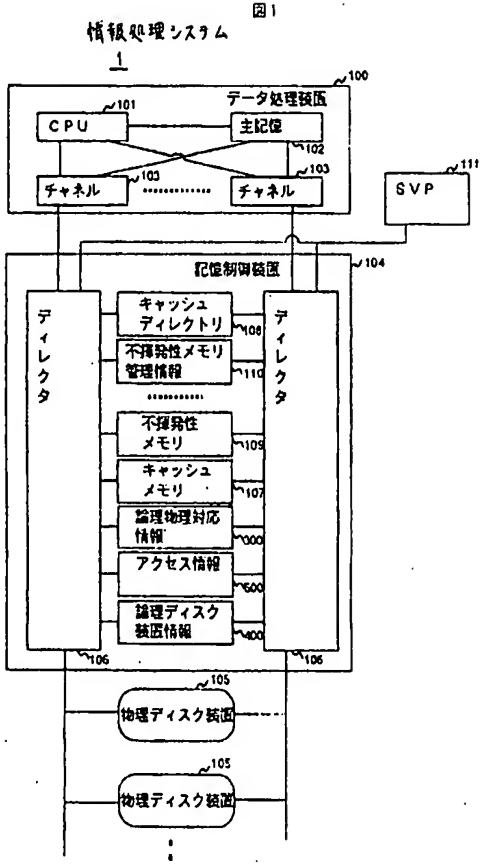
500

論理ディスク装置の数だけ用意

アクセス頻度情報 501
アクセスパターン情報 502

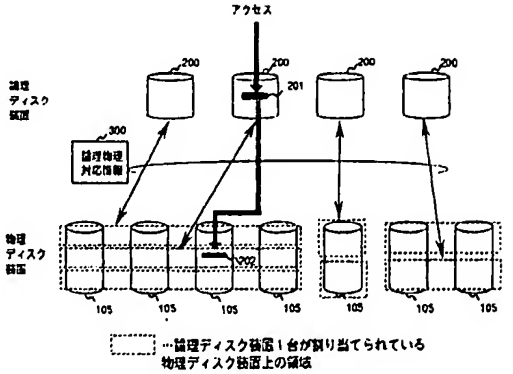
【図1】

図1



【図2】

図2



【図3】

図3

論理物理対応情報
300

・論理ディスク構成情報 310

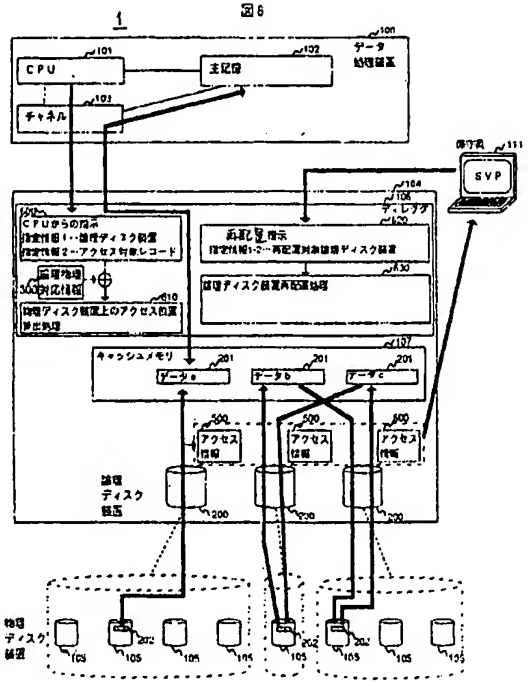
論理ディスク装置の数だけ用意
- 物理ディスク装置グループ 311
- RAID構成 312
- 閉塞位置 313

・物理ディスク構成情報 320

物理ディスク装置の数だけ用意
- 論理ディスク装置グループ 321

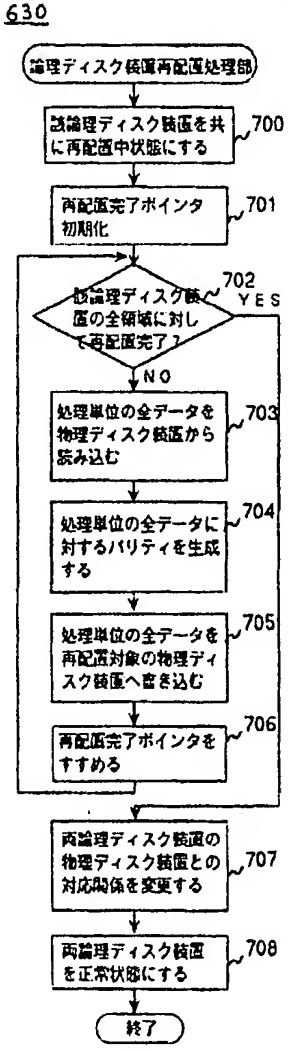
【図6】

図6



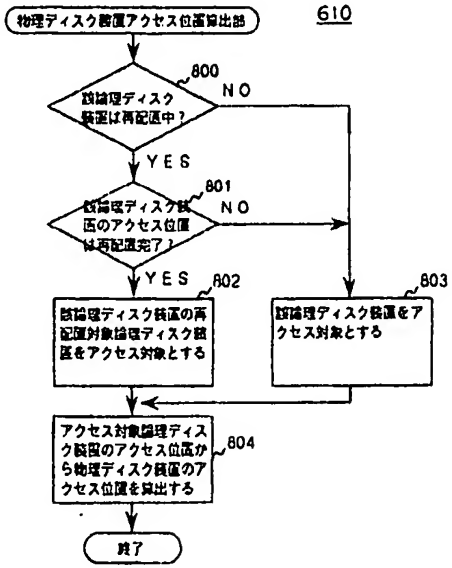
【図7】

図7



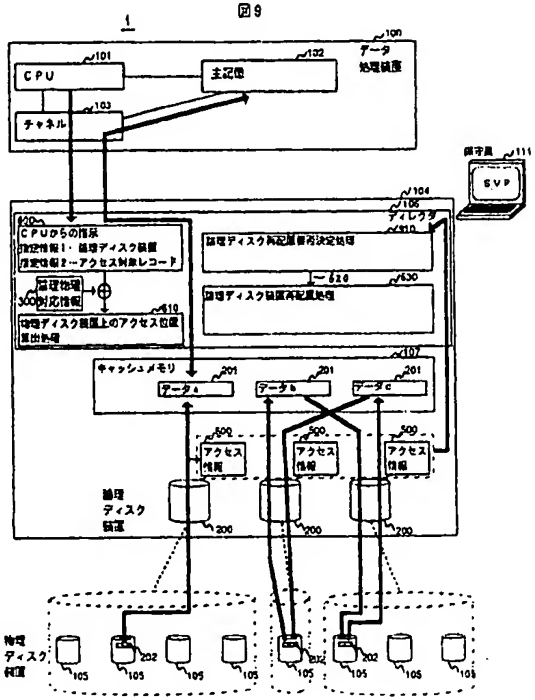
【図8】

図8



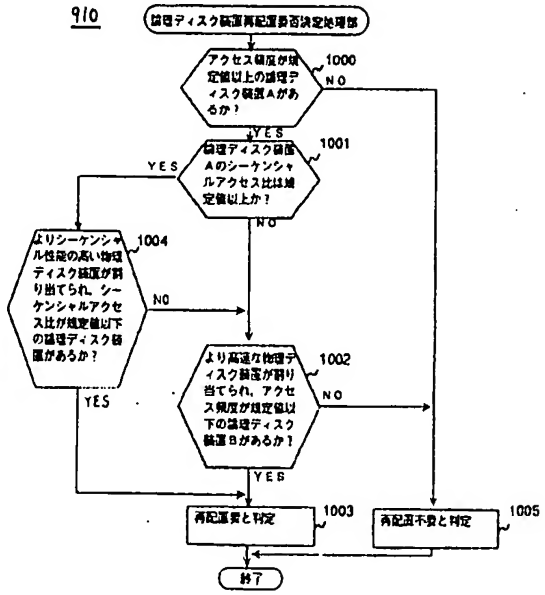
【図9】

図9



【図10】

図10



フロントページの続き

(51)Int.Cl.'	識別記号	FI	ﾃｰﾏｺｰﾄﾞ(参考)
G11B 20/12		G11B 20/12	